

Chapitre 3 – Statistique inférentielle

BTS MEC2 – 2^e année • Version élève

Table des matières

Activités	2
1 Échantillonnage – loi de la moyenne et de la fréquence	2
2 Estimation – ponctuelle et par intervalle de confiance	3
3 Test d’hypothèse – moyenne	5
4 Test d’hypothèse – proportion	7
5 Comparaison de deux échantillons	9
6 Ouverture – Maîtrise statistique des procédés (MSP)	10
7 Bilan – tableau de synthèse	10
8 Exercice de synthèse	10
Carte mentale – vue d’ensemble	11

PROGRAMME BO – BTS ÉCONOMIE DE LA CONSTRUCTION

Contenus : Échantillonnage : loi de la moyenne empirique \bar{X} et de la fréquence empirique F d’un grand échantillon. Simulation d’échantillons. Estimation ponctuelle et par intervalle de confiance (moyenne, proportion). Tests d’hypothèse bilatéral et unilatéral sur une moyenne et une proportion; p -valeur, erreurs de 1^{re} et 2^e espèce, puissance. Comparaison de deux moyennes, de deux fréquences. Lien avec la MSP et les cartes de contrôle.

Démonstrations : Aucune démonstration n’est exigible. Les démonstrations présentées ici (IC d’une moyenne, IC d’une proportion, test bilatéral, protocole p -valeur) sont données à titre de *culture* et pour la cohérence interne du cours.

Capacités : Reconnaître une situation d’échantillonnage et en donner la loi limite. Construire un intervalle de confiance pour une moyenne ou une proportion, interpréter sa signification. Conduire un test d’hypothèse bilatéral ou unilatéral, calculer et interpréter une p -valeur. Comparer deux moyennes ou deux fréquences à partir d’échantillons indépendants. Relier l’inférence statistique aux outils MSP utilisés en entreprise (cartes de contrôle, capacité).

Tout le cours



Loi de la moyenne empirique \bar{X}

Soit X un caractère de moyenne μ et d'écart-type σ sur une population. Pour un échantillon aléatoire simple de taille n suffisamment grande ($n \geq 30$) :

$$\bar{X} \sim \mathcal{N}\left(\mu; \frac{\sigma}{\sqrt{n}}\right).$$

Si X suit déjà une loi normale, ce résultat est exact pour tout $n \geq 1$.

Loi de la fréquence empirique F

Si la proportion d'un caractère dans la population est p , la fréquence F observée sur un échantillon de taille n (avec $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$) vérifie :

$$F \sim \mathcal{N}\left(p; \sqrt{\frac{p(1-p)}{n}}\right).$$

Ces deux résultats sont des conséquences directes du **théorème central limite** (TCL). L'écart-type $\frac{\sigma}{\sqrt{n}}$ (resp. $\sqrt{p(1-p)/n}$) mesure la *précision d'échantillonnage*. Il décroît en $1/\sqrt{n}$: diviser l'erreur par 2 impose de multiplier n par 4.

Simulation d'échantillons – visualisation du TCL

On simule $N = 500$ échantillons de taille $n = 30$ d'une population de moyenne μ . Pour chaque échantillon, on calcule la moyenne \bar{x} . L'histogramme des 500 valeurs obtenues épouse la courbe de la loi $\mathcal{N}(\mu; \sigma/\sqrt{n})$.

30

Plus n augmente, plus l'histogramme se *resserre* autour de μ . C'est la version expérimentale de la loi faible des grands nombres : $\bar{X} \rightarrow \mu$ en probabilité. Cette simulation (Python, tableur, calculatrice) est la meilleure image de l'inférence statistique : on observe *une* valeur \bar{x} , extraite d'une distribution qu'on ne voit jamais entièrement.

On sait que $X \sim \mathcal{N}(10; 0,4)$ mm. Pour $n = 16$, donner la loi de \bar{X} puis calculer $P(9,8 \leq \bar{X} \leq 10,2)$.

2 Estimation – ponctuelle et par intervalle de confiance

Estimation ponctuelle

Soit un échantillon de taille n issu d'une population :

$$\hat{\mu} = \bar{x}, \quad \hat{p} = f, \quad \hat{\sigma}^2 = \frac{n}{n-1} s^2.$$

\bar{x} estime la moyenne, f estime la proportion, $\frac{n}{n-1} s^2$ estime la variance (*correction sans biais* de Bessel ; s^2 est la variance empirique de l'échantillon).

Intervalle de confiance d'une moyenne (n grand)

Au niveau de confiance $1 - \alpha$:

$$\text{IC}_{1-\alpha}(\mu) = \left[\bar{x} - z_{\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}} ; \bar{x} + z_{\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}} \right].$$

Valeurs usuelles : $z_{0,025} = 1,96$ (95 %) ; $z_{0,005} = 2,576$ (99 %).

Sous les hypothèses du TCL, $\bar{X} \sim \mathcal{N}(\mu; \sigma/\sqrt{n})$ donc la variable centrée réduite $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ suit $\mathcal{N}(0, 1)$.

Par définition de $z_{\alpha/2}$:

$$P\left(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1 - \alpha.$$

On isole μ (inégalités équivalentes) :

$$P\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

Appliqué à l'échantillon observé, avec $\hat{\sigma}$ à la place de σ inconnu, on obtient l'IC annoncé.

Intervalle de confiance d'une proportion

$$\text{IC}_{1-\alpha}(p) = \left[f - z_{\alpha/2} \sqrt{\frac{f(1-f)}{n}} ; f + z_{\alpha/2} \sqrt{\frac{f(1-f)}{n}} \right].$$

D'après le TCL, $F \sim \mathcal{N}(p; \sqrt{p(1-p)/n})$. On pose $Z = \frac{F - p}{\sqrt{p(1-p)/n}} \sim \mathcal{N}(0, 1)$, d'où

$$P\left(|F - p| \leq z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha.$$

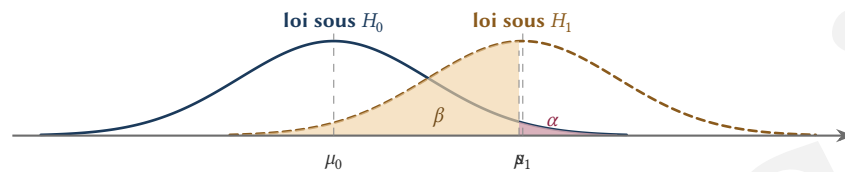
La variance $p(1-p)$ étant inconnue, on la remplace par son estimation $f(1-f)$ (approximation valide pour n grand) : on obtient l'IC d'une proportion.

Risques d'erreur, p -valeur et puissance

Deux décisions, deux erreurs possibles :

	H_0 vraie	H_0 fausse
On rejette H_0	erreur α	décision correcte (puissance $1 - \beta$)
On ne rejette pas	décision correcte	erreur β

- α : risque de rejeter H_0 à tort (1^{re} espèce); c'est le seuil choisi.
- β : risque d'accepter H_0 à tort (2^e espèce); dépend de la vraie valeur $\mu_1 \neq \mu_0$.
- **Puissance** $\pi(\mu_1) = 1 - \beta(\mu_1)$: probabilité de détecter un écart quand il existe.
- **p -valeur** : plus petit seuil α pour lequel on rejetterait H_0 . Plus p est petite, plus les données sont incompatibles avec H_0 .



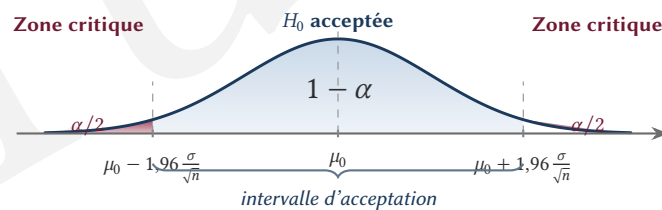
Test bilatéral d'une moyenne

$H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$.

Sous $H_0 : \bar{X} \sim \mathcal{N}\left(\mu_0; \frac{\sigma}{\sqrt{n}}\right)$.

Zone d'acceptation : $\left[\mu_0 - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}; \mu_0 + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$.

p -valeur : $p = 2 P\left(Z \geq \left| \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \right| \right)$.



Sous $H_0, Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$. On cherche $W_\alpha = \{|Z| > c\}$ tel que $P(|Z| > c) = \alpha$, soit $c = z_{\alpha/2}$.

La zone de rejet sur \bar{X} se déduit par :

$$|Z| > z_{\alpha/2} \iff |\bar{X} - \mu_0| > z_{\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

D'où la zone d'acceptation symétrique autour de μ_0 . Le complément de α (probabilité d'être dans W_α sous H_0) est réparti à parts égales dans les deux queues.

Test unilatéral d'une moyenne

$H_0 : \mu = \mu_0$ contre $H_1 : \mu < \mu_0$ (ou $\mu > \mu_0$).

Zone de rejet unilatérale gauche : $\bar{x} < \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}}$.

Valeurs usuelles : $z_{0,05} = 1,645$; $z_{0,01} = 2,326$.

Lorsque les conditions d'approximation normale ne sont pas remplies ($n < 30$, ou $np_0 < 5$, ou $n(1 - p_0) < 5$), on revient à la loi binomiale exacte sous $H_0 : X \sim \mathcal{B}(n, p_0)$.

Test unilatéral à droite ($H_1 : p > p_0$) : on rejette H_0 si la p -valeur $P(X \geq x_{\text{obs}}) \leq \alpha$ (calcul direct par calculatrice : $1 - \text{binomCdf}(n, p_0, x_{\text{obs}} - 1)$).

Test bilatéral : $p\text{-valeur} = 2 \min(P(X \leq x_{\text{obs}}), P(X \geq x_{\text{obs}}))$.

Exemple : sur $n = 15$ pièces, on en trouve 4 défectueuses alors que $p_0 = 0,10$.

$P(X \geq 4) = 1 - P(X \leq 3)$ avec $X \sim \mathcal{B}(15; 0,10)$, soit environ $0,055$: on ne rejette pas H_0 au seuil 5% .

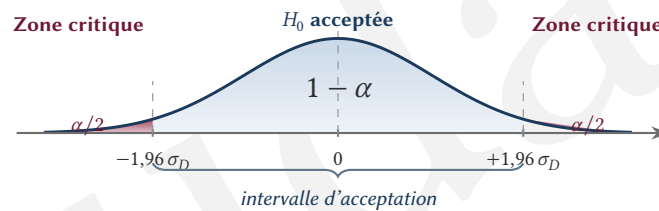
5 Comparaison de deux échantillons

Comparaison de deux moyennes

Soient deux échantillons indépendants, $H_0 : \mu_1 = \mu_2$. Sous H_0 :

$$D = \bar{X}_1 - \bar{X}_2 \sim \mathcal{N}\left(0; \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right).$$

Zone d'acceptation bilatérale : $|d| \leq z_{\alpha/2} \sigma_D$.



Comparaison de deux fréquences

Sous $H_0 : p_1 = p_2 = p$, on estime la proportion commune

$$\hat{p} = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2} \quad \text{puis} \quad \sigma_D = \sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}.$$

Zone d'acceptation : $|f_1 - f_2| \leq z_{\alpha/2} \sigma_D$.

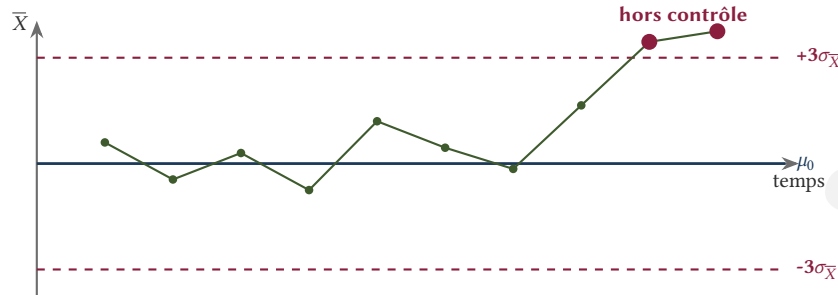
Sacs A : $n_1 = 50$, $\bar{x}_1 = 72,8$ g, $\sigma_1 = 3$.

Sacs B : $n_2 = 50$, $\bar{x}_2 = 74,2$ g, $\sigma_2 = 3,2$.

Différence significative à 5% ?

6 Ouverture – Maîtrise statistique des procédés (MSP)

En entreprise, l'inférence statistique sert à *piloter* un procédé de fabrication : c'est la **MSP** (Maîtrise Statistique des Procédés). Une **carte de contrôle** est la mise en œuvre graphique d'un test d'hypothèse répété : on prélève régulièrement un échantillon, on calcule \bar{X} et on le reporte sur un graphique où figurent la cible μ_0 et deux *limites de contrôle* $\mu_0 \pm 3\sigma/\sqrt{n}$ (règle des 3 σ , soit un risque $\alpha \approx 0,27\%$).



La MSP fait le pont avec les enseignements techniques : en **qualité**, on parle de *capabilité* $C_p = \frac{IT}{6\sigma}$ et d'*indice d'aptitude*. En **production**, tout dérèglement (point hors limites, suite de points du même côté) déclenche une intervention sur la machine. Le *vocabulaire* de l'inférence (hypothèse, seuil, rejet, *p*-valeur, puissance) est ainsi employé quotidiennement dans l'industrie.

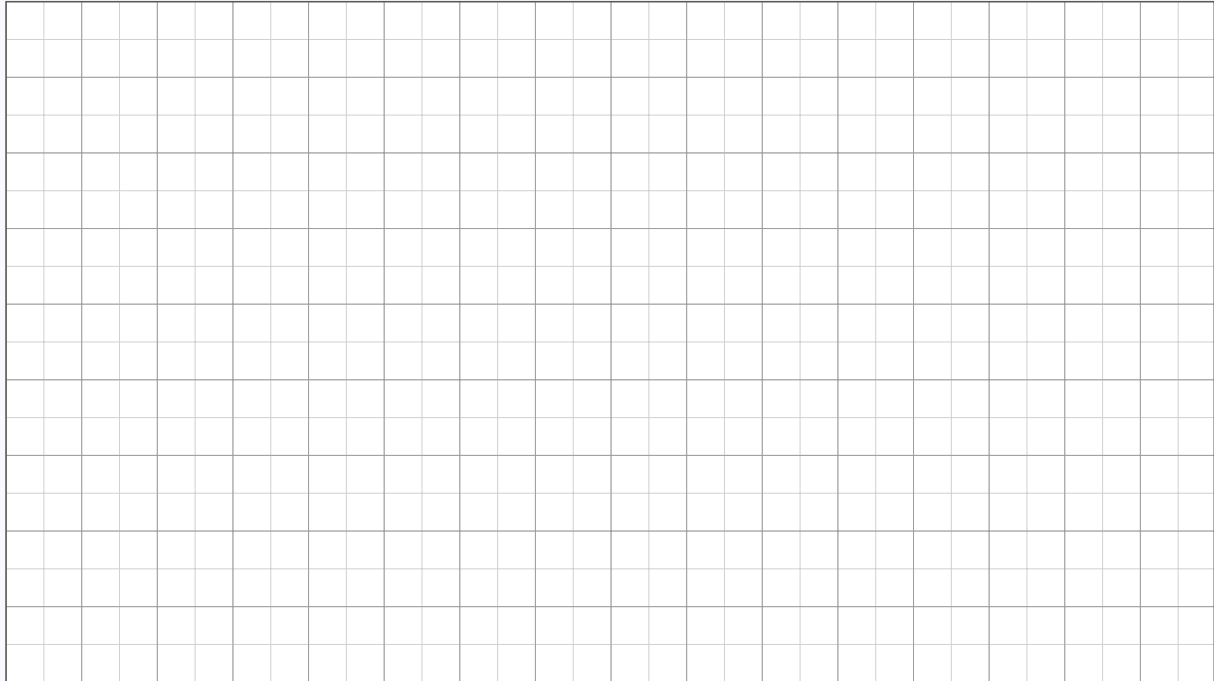
7 Bilan – tableau de synthèse

Situation	Statistique	Loi / formule
Moyenne d'un échantillon	\bar{X}	$\mathcal{N}(\mu; \sigma/\sqrt{n})$
Fréquence d'un échantillon	F	$\mathcal{N}(p; \sqrt{p(1-p)/n})$
IC d'une moyenne	$\bar{x} \pm z_{\alpha/2} \hat{\sigma} / \sqrt{n}$	$z_{0,025} = 1,96$
IC d'une proportion	$f \pm z_{\alpha/2} \sqrt{f(1-f)/n}$	$z_{0,025} = 1,96$
Test moyenne bilatéral	$ \bar{x} - \mu_0 \leq z_{\alpha/2} \sigma / \sqrt{n}$	acceptation
Test moyenne unilatéral	$\bar{x} \leq \mu_0 \pm z_{\alpha} \sigma / \sqrt{n}$	rejet
Test proportion	$ f - p_0 \leq z_{\alpha/2} \sqrt{p_0(1-p_0)/n}$	idem
Comparaison moyennes	2 $D = \bar{X}_1 - \bar{X}_2$	$\mathcal{N}(0; \sigma_D)$
<i>p</i> -valeur (bilat.)	$p = 2 P(Z \geq z_{\text{obs}})$	rejet si $p < \alpha$
Puissance	$1 - \beta = P(\text{rejeter} H_1)$	↑ avec n et $ \mu_1 - \mu_0 $

8 Exercice de synthèse

Une usine produit des bouteilles dont le volume X suit $\mathcal{N}(\mu; 2)$ mL. La norme impose $\mu = 500$ mL. Un contrôle sur $n = 64$ bouteilles donne $\bar{x} = 499,3$ mL.

1. Donner la loi de \bar{X} sous la norme.
2. Construire un intervalle de confiance de μ à 95 %.
3. Tester $H_0 : \mu = 500$ au seuil 1 % (bilatéral).
4. Calculer la p -valeur et la comparer au seuil.
5. La norme est-elle respectée ?



Carte mentale – vue d’ensemble

